Supplementary Figures for:

Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology

Elior Rahmani[1], Regev Schweiger[2], Brooke Rhead[3], Lindsey A. Criswell[4], Lisa F. Barcellos[3], Eleazar Eskin[1,5], Saharon Rosset[6], Sriram Sankararaman[1], Eran Halperin[1,5,7]

[1]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA

[2]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

[3]School of Public Health, University of California, Berkeley, Berkeley, CA, USA

[4]Russell / Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

[5]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

[6]Department of Statistics, Tel Aviv University, Tel Aviv, Israel

[7]Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA, USA

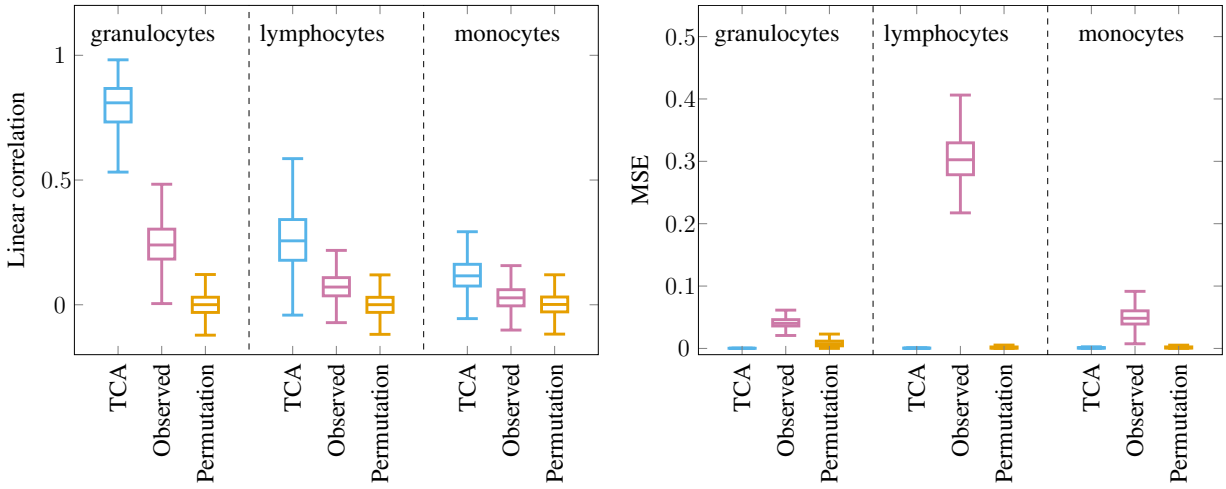

Figure S1: Reconstructing cell-type-specific methylation levels from simulated bulk whole-blood data with three constituting cell types ($k = 3$; 250 samples, 250 sites). Three approaches were evaluated in capturing the cell-type-specific levels of each site j and cell type h across all individuals $z_{hj} = (z_{hj}^1, ..., z_{hj}^n)$: TCA, TCA after permuting the observed data matrix ("Permutation") and directly using the observed bulk data ("Observed"; i.e. using the bulk as the estimate for the cell-type-specific levels of each cell type). For each of the evaluated approaches and for each of the simulated cell types (ordered by their mean abundance), presented are the distributions of the linear correlation between z_{hj} and its estimate $\hat{z}_{hj}$ across all sites j and across ten simulated data sets (left), and the distribution of the MSE between z_{hj} and its estimate $\hat{z}_{hj}$ across all sites j and across ten simulated data set (right).

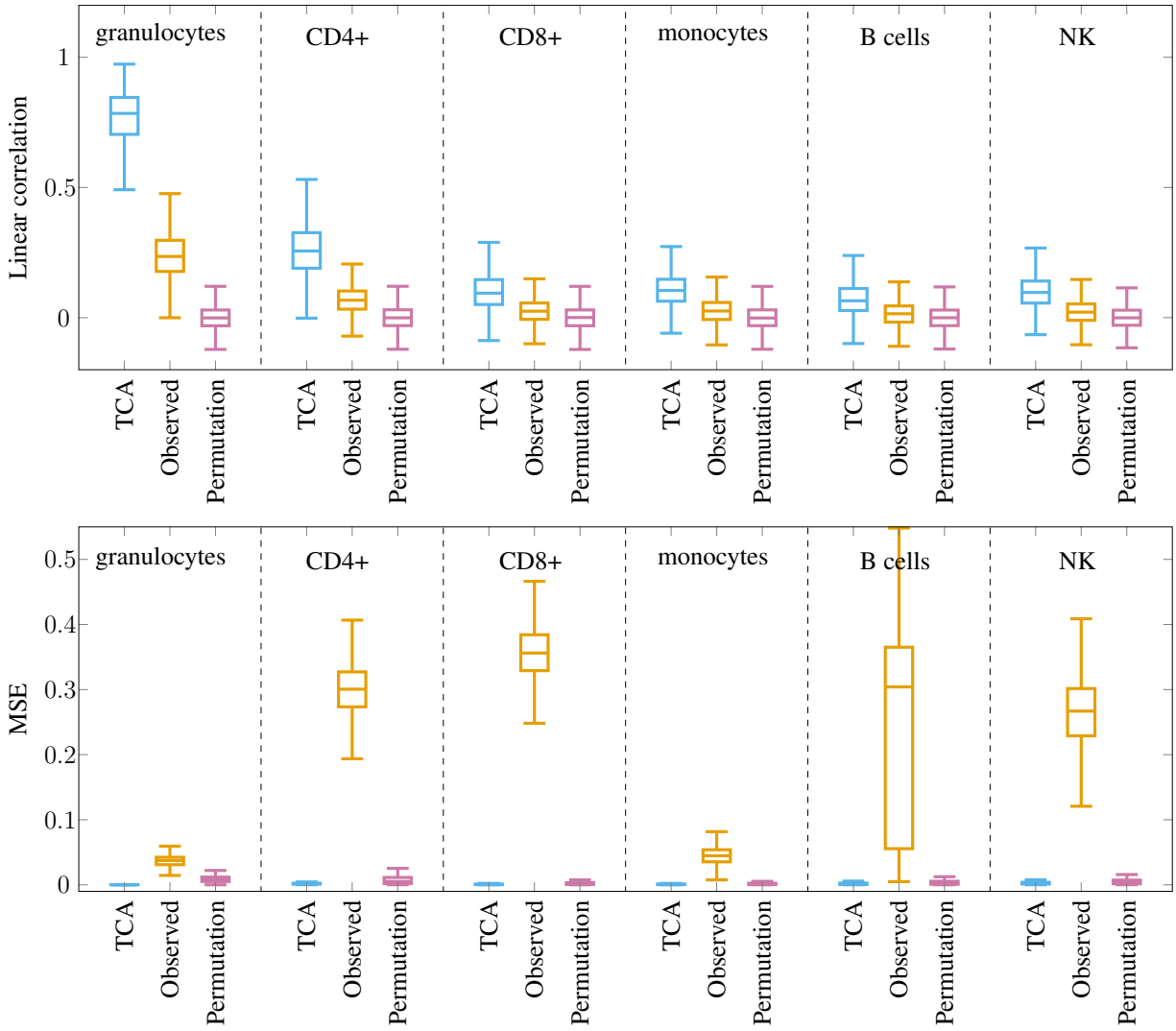

Figure S2: Reconstructing cell-type-specific methylation levels from simulated bulk whole-blood data with six constituting cell types ($k = 6$; 250 samples, 250 sites). Three approaches were evaluated in capturing the cell-type-specific levels of each site j and cell type h across all individuals $z_{hj} = (z_{hj}^1, ..., z_{hj}^n)$: TCA, TCA after permuting the observed data matrix ("Permutation") and directly using the observed bulk data ("Observed"; i.e. using the bulk as the estimate for the cell-type-specific levels of each cell type). For each of the evaluated approaches and for each of the simulated cell types (ordered by their mean abundance), presented are the distributions of the linear correlation between z_{hj} and its estimate $\hat{z}_{hj}$ across all sites j and across ten simulated data sets (top), and the distribution of the MSE between z_{hj} and its estimate $\hat{z}_{hj}$ across all sites j and across ten simulated data set (bottom).

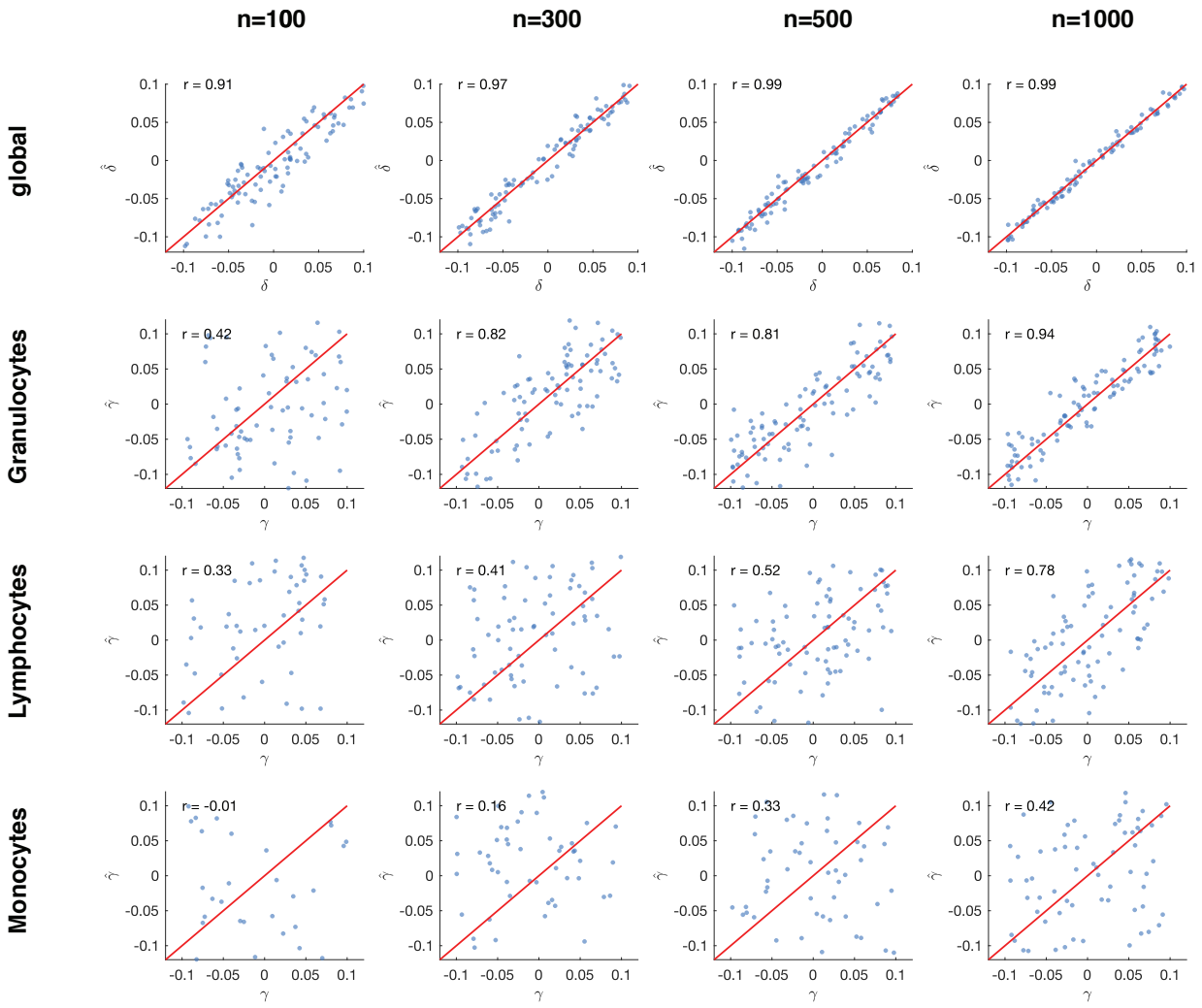

Figure S3: Estimating the effect sizes of covariates affecting methylation using TCA. Presented are true (X axes) and estimated (Y axes) effect sizes in simulated whole-blood methylation data with three constituting cell types ($k = 3$) and varying sample sizes (separated by different columns). Two scenarios were considered using a range of effect sizes: (1) estimating the effect of a covariate with global (i.e. non-cell-type-specific) effect on methylation (top row), and (2) estimating the effect of a covariate with cell-type-specific effect on methylation (rows 2-4). In the latter scenario, we considered a separate experiment for each of the three cell types, such that in each experiment the covariate affected the particular cell type under test. Throughout these experiments, covariates were generated from a normal distribution, and both global (δ) and cell-type-specific (γ) effect sizes were sampled from a uniform distribution.

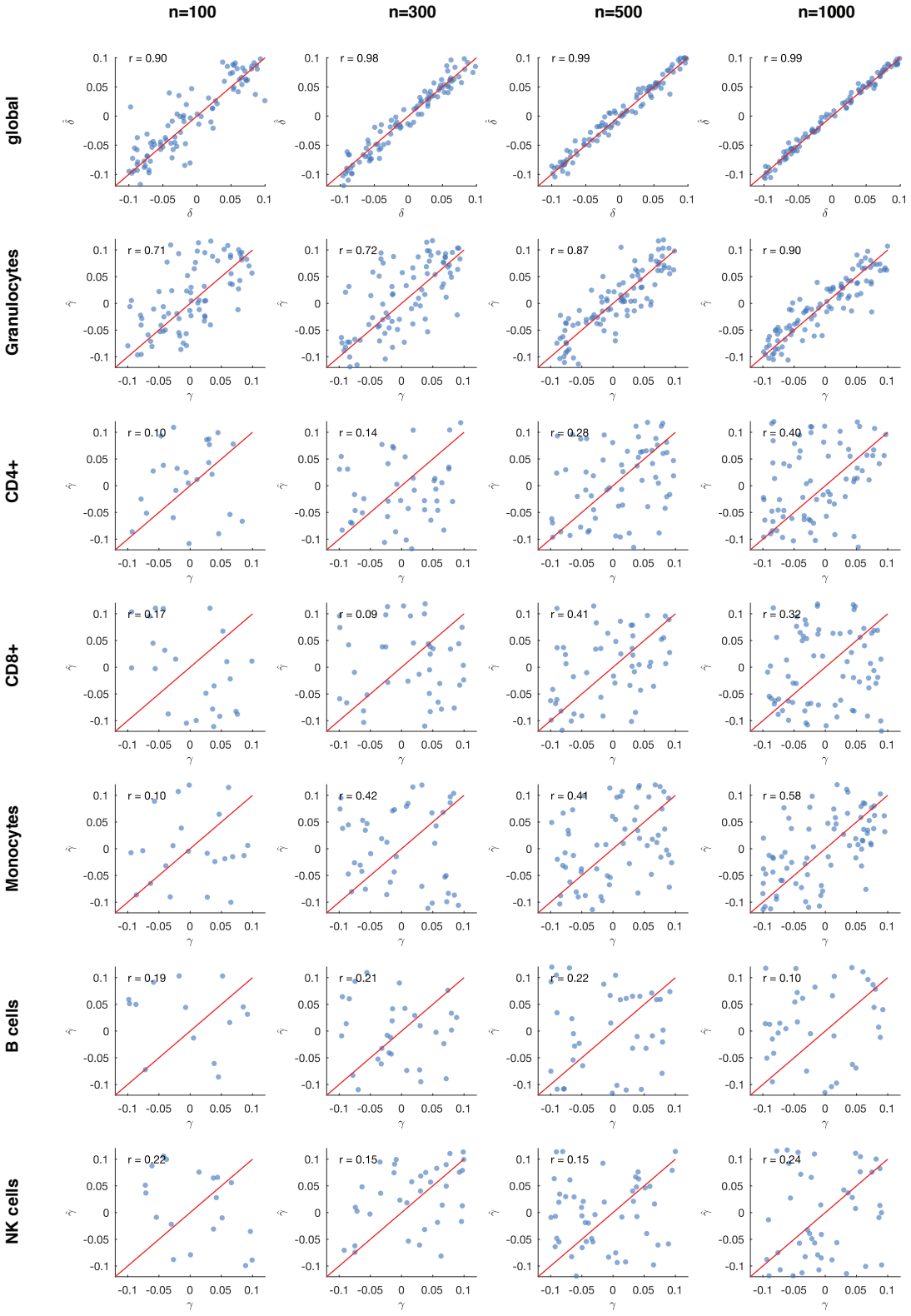
n=100
n=300
n=500
n=1000
global
Granulocytes
CD4+
CD8+
Monocytes
B cells
NK cells
r = 0.90
r = 0.98
r = 0.99
r = 0.99
r = 0.71
r = 0.72
r = 0.87
r = 0.90
r = 0.10
r = 0.14
r = 0.28
r = 0.40
r = 0.17
r = 0.09
r = 0.41
r = 0.32
r = 0.10
r = 0.42
r = 0.41
r = 0.58
r = 0.19
r = 0.21
r = 0.22
r = 0.10
r = 0.22
r = 0.15
r = 0.15
r = 0.24

Figure S4: Estimating the effect sizes of covariates affecting methylation using TCA. Presented are true (X axes) and estimated (Y axes) effect sizes in simulated whole-blood methylation data with six constituting cell types ($k = 6$) and varying sample sizes (separated by different columns). Two scenarios were considered using a range of effect sizes: (1) estimating the effect of a covariate with global (i.e. non-cell-type-specific) effect on methylation (top row), and (2) estimating the effect of a covariate with cell-type-specific effect on methylation (rows 2-7). In the latter scenario, we considered a separate experiment for each of the three cell types, such that in each experiment the covariate affected the particular cell type under test. Throughout these experiments, covariates were generated from a normal distribution, and both global (δ) and cell-type-specific (γ) effect sizes were sampled from a uniform distribution.

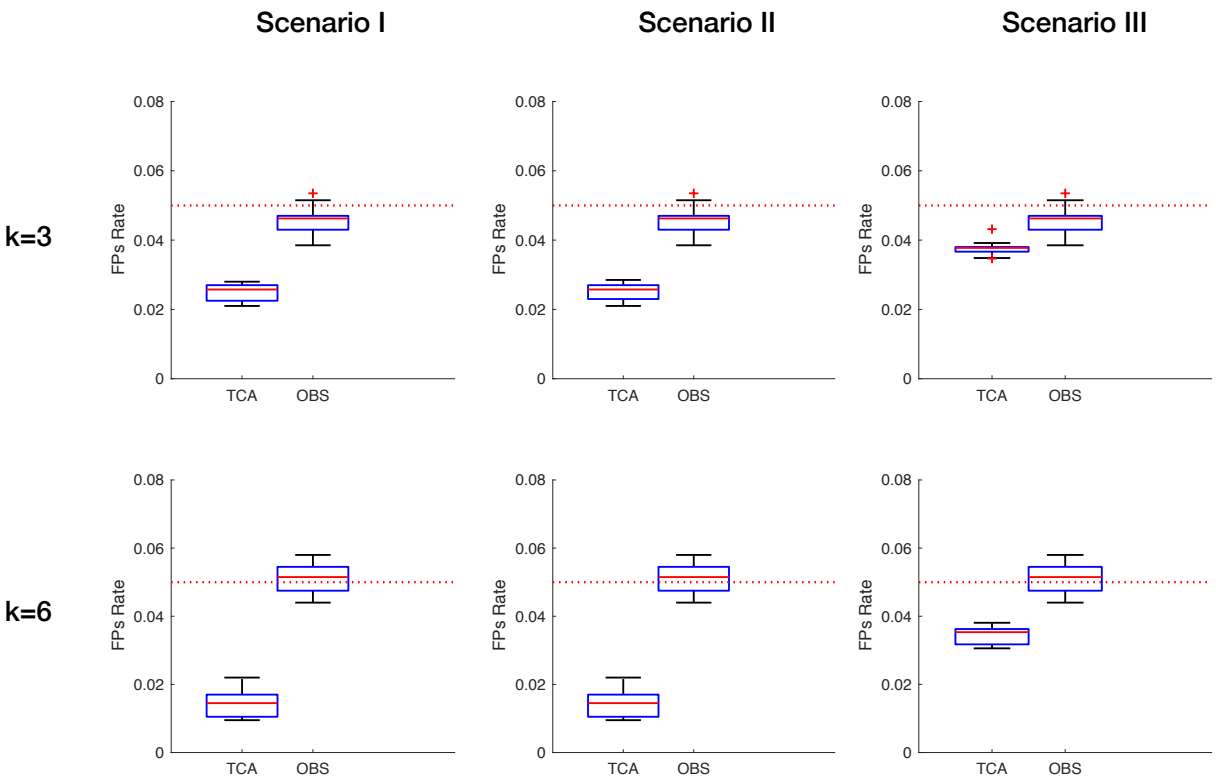

Figure S5: An evaluation of false positives rates in association testing with DNA methylation. Performance was evaluated using two approaches: TCA and a standard linear regression with the observed bulk data (OBS). The proportions of false positives (FPs) were measured under three scenarios using a range of effect sizes: different effect sizes for different cell types (Scenario I), the same effect size for all cell types (Scenario II), and only a single effect size for a single cell type (Scenario III); each of the scenarios was evaluated under the assumption of three constituting cell types (k=3) and six constituting cell types (k=6). Boxplots reflect results across multiple simulations.

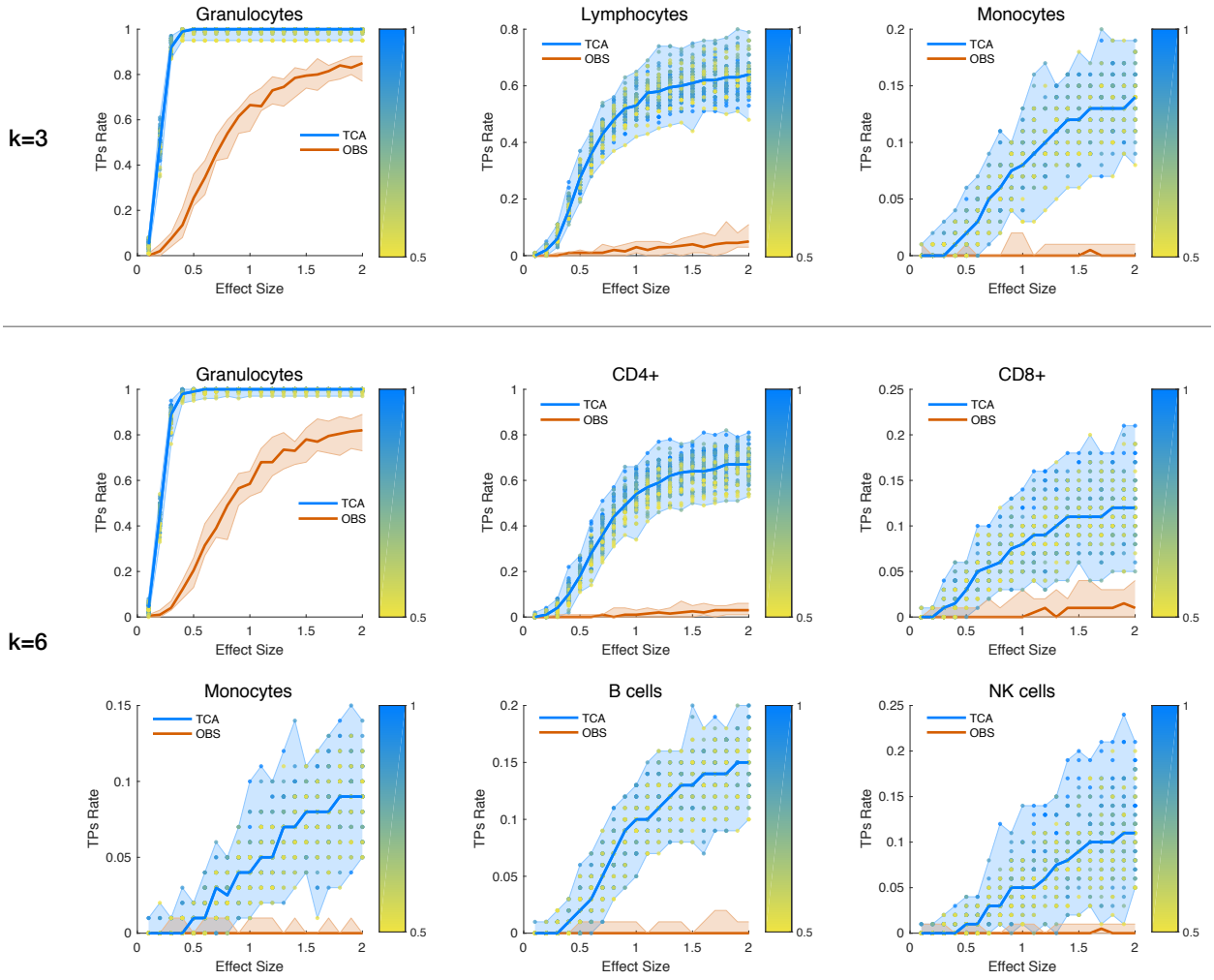

Figure S6: An evaluation of power for detecting cell-type-specific associations with DNA methylation, stratified by cell types. Performance was evaluated using two approaches: TCA with a marginal test and a standard linear regression with the observed bulk data (OBS). The numbers of true positives were measured under a scenario where only a single effect size for a single cell type exists, both in the case of three constituting cell types (k=3) and six constituting cell types (k=6). The colored areas reflect results across multiple simulations, and the colored dots reflect the results of TCA under different initializations of the cell-type composition estimates, where the color gradients represent the mean absolute correlation of the initial estimates with the true values (across all cell types).

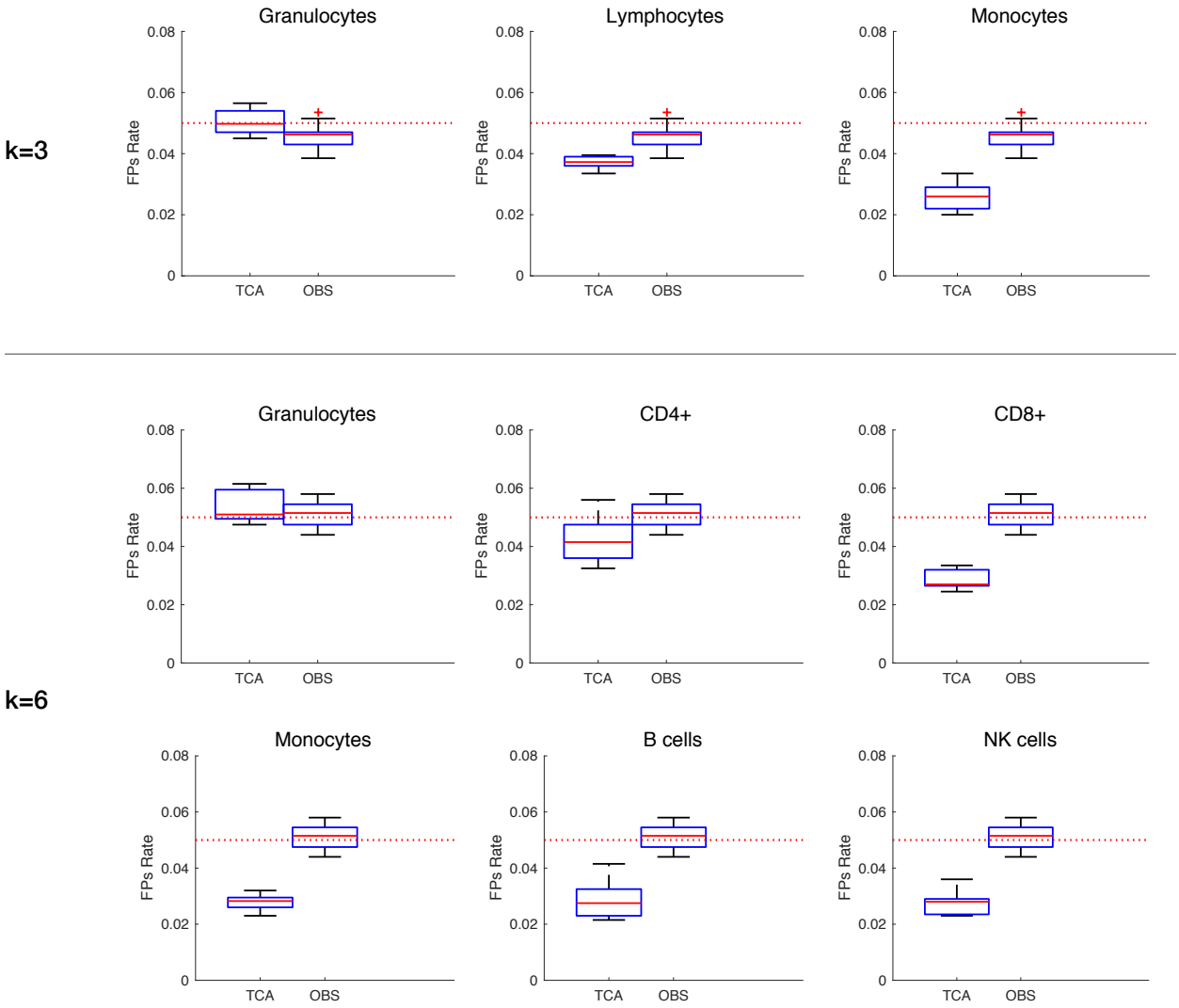

Figure S7: An evaluation of false positives rates in association testing with DNA methylation, stratified by cell types. Performance was evaluated using two approaches: TCA with a marginal test and a standard linear regression with the observed bulk data (OBS). The proportions of false positives (FPs) were measured under a scenario where only a single effect size for a single cell type exists, both in the case of three constituting cell types (k=3) and six constituting cell types (k=6). Boxplots reflect results across multiple simulations.